Dear Reviewer,

Thank you for your valuable comments and suggestions. Below, we would like to address your concerns point by point.

Comment 1:

Spatial distribution of crops is critical information for food security, agriculture development and investment decisions, sustainable agricultural development etc. There have been multiple attempts, by different teams in the world, to produce global crop maps. And yet so far few attempts have been made to produce time series global crop maps. The GGCP10 dataset focuses on maize, wheat, rice, and soybeans and covers the years 2010 to 2020, the first temporally continuous, gridded dataset of crop production at the global scale. The dataset was constructed using a data-driven spatial production allocation model that incorporated multiple source datasets. The use of various data sources, including FAO statistical data, GAEZ+ 2015 annual crop data, and other sources, demonstrates a robust foundation for the study. This model was rigorously examined through pre-processing and consistency checks to ensure data accuracy and reliability. The incorporation of machine learning techniques for predicting crop yields and production is a forward-looking approach. These techniques have demonstrated solid performance in recent years. The approach of combining information from multiple sources, including climate, soil, and topographic data, is a commendable strategy for predicting crop production accurately.

Response to Comment 1:

We sincerely appreciate your positive evaluation of our work and your insightful summary of the significance of the paper. Your high appraisal of the importance of this research in constructing a temporally continuous and globally spatially covered gridded crop production dataset, as well as the robustness and innovativeness of the data-driven spatial production allocation model in integrating multi-source datasets, will be a great encouragement for our future research work. We will further refine the paper according to your valuable comments, striving to provide more long-term and accurate data support for global agricultural production mapping and food security research. Thank you again for your detailed review and valuable suggestions. We cordially invite you to continue providing us with your insightful feedback and guidance when we submit the revised manuscript.

Comment 2:

My first concern is that their whole modelling approach, production model in particular (see Section 2.2.3 Data-driven Model Training), implicitly assumes that the biophysical parameters alone could determine the crop production. In other words, their modelling approach assumes that the driven factors for the huge spatial heterogeneity of crop productivity (or production if crop area is counted) are mainly those biophysical parameters such as soil, AEZ zones, various vegetation indices, climate variables (multi-source indicators XI(i,j) as shown in their model, Line 225). Any breeders or agricultural economists would tell you that this is not true. Social economic factors such as crop seeds/varieties, crop management, fertilizer, pesticide are the major driven force in crop productivity (and so crop production). This is why, for example, the maize yield in a large estate farm in Zambia could be a few times higher than that of a subsistence maize farmer next door – just a few hundred meters away! Of course collecting the data for these parameters on a global scale is much harder, if possible at all. Without the inputs of these critical parameters, estimating crop yields spatially is a huge challenge.

Response to Comment 2:

We appreciate your insightful comments. We agree that socio-economic factors (such as crop management practices) play a crucial role in shaping the spatial heterogeneity of crop production. Obtaining such data on a global scale poses significant challenges, which is one of the limitations of the current research. We will supplement the discussion section of the paper to elaborate on this limitation.

1) It is worth noting that although we lack direct data on crop management practices, **some of the factors included in the model can, to a certain extent, reflect the spatial differences in crop management levels.** For example, irrigation data reflects differences in irrigation management inputs, which greatly influence crop growth and production. Crop planting area data partially reflects farmers' planting preferences and resource allocation decisions for different crops, demonstrating farmers' responses to market conditions and policies.

2) Some of the remote sensing-derived indicators we included can also reflect the impact of crop management to a certain degree. For instance, the Maximum Vegetation Condition Index (VCIx) describes the historical relative level of vegetation conditions during the study period. A higher VCIx indicates relatively better crop growth during that period, which to some extent benefits from farmers' good field management. Indicators such as Net Primary Productivity (NPP) and Leaf Area Index (LAI) reflect crop biomass accumulation and photosynthetic intensity. Higher NPP and LAI are often the result of good management.

3) Moreover, agro-ecological zone data comprehensively considers the impact of natural conditions such as climate, soil, and topography on agricultural production. Different ecological conditions often correspond to differentiated planting systems

and management patterns. Therefore, one important reason we chose to model at the agro-ecological zone scale is that we hope the model can characterize the spatial variation of production within the ecological zone through the differences in multiple variables within the zone.

4) Of course, we also recognize that due to data limitations, the current model does not incorporate key crop management factors such as cropping systems, variety selection, and fertilizer use, which may obscure some important drivers of production variation. Therefore, in the discussion section, we will further analyze this limitation of the model and its impact on data quality.

Comment 3:

My second concern is that the paper is a data description paper and yet it misses the critical dataset: a global sub-national crop statistics data. Their major statistical data source is the FAOSTATA data at country level, which is too coarse for the gridded product. Crop type mapping is too complex and too dynamic to be able to be modeled without the actual sub-national statistics. For example, farmers may decide to reduce their maize area and instead plant more rice in the current season if they expect more rain in the coming season or simply they believe the maize price will go down next year. Any fancy modelling approach is difficult to capture that without the actual data. The paper itself emphasizes a lot on their modelling approach while ignoring the time-consuming effort of collecting crop data for the four crops (maize, rice, wheat and soybean). I would say the latter is much more critical, in particular considering that the ESSD journal is, which I quote, "for the publication of articles on original research data (sets), furthering the reuse of high-quality data of benefit to Earth system sciences".

Response to Comment 3:

1) We strongly agree with your view on the importance of crop statistics at the national/regional level. However, we would like to further clarify that in this study, **national-level statistics are mainly used for post-processing of model results**, i.e., calibrating the gridded production estimates with national official statistics through consistency processing to ensure statistical consistency of the estimates at the administrative unit level. During the model training and prediction stages, we mainly use spatialized multi-source data such as remote sensing-derived indicators, meteorological and soil data, etc., which can provide high-resolution crop growth information to support fine-scale production mapping. Therefore, the spatial resolution of statistical data does not directly affect the modeling accuracy.

2) Indeed, obtaining crop statistics at sub-national administrative levels is of great importance for understanding farmers' planting behaviors and assessing the impact of regional agricultural policies. To this end, we have made every effort to collect state/provincial-level crop statistics from major agricultural countries such as the United States, Canada, Argentina, Brazil, India, China, Thailand, and Australia

through official channels and partnerships. However, as the reviewer mentioned, these data vary greatly in terms of access channels, update frequency, and spatiotemporal coverage, making it difficult to fully unify the format and connotation of data from different countries. If they are used for consistency calibration, it may be difficult to ensure statistical consistency at the national scale.

3) In contrast, **the national agricultural statistics released by the FAO have obvious advantages**. Although the spatial resolution is coarser, the data sources are authoritative, the time series is long, and the official statistics of various countries have been systematically summarized and verified. At the current stage, using FAO data for global-scale production estimation result calibration can maximize the use of existing data resources while ensuring the consistency of data benchmarks across countries. Such global production mapping data based on a unified benchmark can better serve applications such as monitoring the SDGs and assessing global food security. Of course, in the long run, agricultural statistics at sub-national administrative levels are irreplaceable for understanding regional agricultural production processes and optimizing resource allocation.

4) In addition, we would like to further emphasize that the collection, production and processing of the large amount of basic data used in this study is very time-consuming. In addition to using publicly published data products, the potential biomass, CALF, and VCIx indicators we use require the research team to go through a series of complex processing steps from raw data acquisition to final indicator generation, including data storage, format conversion, radiometric calibration, atmospheric correction, geometric correction, cloud and snow masking, vegetation index calculation, and pixel compositing. Each step requires professional algorithm design and parameter tuning to ensure data quality.

Comment 4:

The paper could benefit from more transparency regarding data preprocessing steps, such as how data clipping based on crop phenology is conducted and how missing or corrupted data are handled. For example, Line 179-183, Where does CA(i,t) come from? How to divide CA(i, t) into CA(i,j) ? Not clear at al. I think (I am not 100% sure as I have a hard time to understand this section) "reference year" at Line 184 should be "target year". After reading the section multiple times, I still don't know how the harvested area is estimated at the pixel level. I considered myself as an expert, imagine how an ordinary reader would feel!

Response to Comment 4:

Thank you for your thorough review and constructive suggestions. We will follow your advice to provide a clearer and more detailed description of the data preprocessing steps in the data and methods section, so that readers can accurately understand each processing step.

1) First, **regarding data clipping based on crop phenology, we adopt a method of extracting time windows and calculating feature indicators from time-series data based on crop phenology information**. Specifically, we first determine the time range of the main growth stages (such as sowing, growing and maturity) for each crop type according to the crop phenology; then, we extract the corresponding time-series data based on the main growth stages of each crop as time windows; finally, within each time window, we calculate the statistical feature values (such as maximum, minimum, standard deviation and total sum) of the time-series indicators (such as cumulative precipitation, average air temperature) and use them as input features for the model. For the processing of missing or abnormal data, we will also provide detailed technical details in the revised manuscript to facilitate readers' understanding and improve the transparency and reproducibility of the data processing methods.

2) Secondly, **regarding the lack of clarity in the description of harvested area estimation** at the pixel level in lines 179-183 and the surrounding context. The estimation of harvested area is one of the important innovations in this study. We combine the gridded data (harvested area, planted area) of the reference year, the gridded data (planted area) and statistical data (harvested area) of the target year to dynamically estimate the harvested area of each grid within the region. Due to the limited space, we will introduce the principles and processes of harvested area estimation in detail in the revised manuscript, striving to clearly describe this innovative method so that both experts and ordinary readers can better understand it.

3) **Regarding other detailed issues you mentioned**, such as CA(i,t), CA(i,t) represents the total planted area of all crops in the i-th grid in the t-th growing season, which is obtained by multiplying the cropped arable land fraction (CALF) of the i-th grid by the cropland area of that grid. For line 184, "reference year" should indeed be "target year". We will carefully proofread and refine this part of the content in the revised manuscript.

Thank you again for your valuable comments. We will revise the paper accordingly to meet the publication requirements.

Comment 5:

Model Selection: The paper mentions the selection of machine learning models but lacks specific details about the criteria used for model selection. Providing more insight into the model selection process would enhance the paper's transparency.

Response to Comment 5:

We greatly appreciate your suggestion. We agree that providing more details in the model selection section will help improve the transparency of the research. We will expand and refine this part of the content. Specific revisions include:

1) First, in the revised manuscript, **we will explain in detail the reasons for choosing machine learning models**. Compared to traditional statistical models, machine learning models have advantages in dealing with complex nonlinear relationships and high-dimensional data. At the same time, our previous research has shown that machine learning models perform well in crop production estimation[1]. Therefore, we adopt machine learning models to construct the spatial allocation relationship of production. Regarding the selection of Random Forest, XGBoost, and CatBoost as candidate models, it is mainly based on their good performance and robustness in the field of geospatial modeling: Random Forest is one of the most commonly used algorithms in ensemble learning; XGBoost has achieved excellent results in various data mining competitions and has been proven to have significant advantages in processing high-dimensional and nonlinear relationship data; CatBoost has shown outstanding performance in multiple data science competitions and is considered a powerful tool for handling mixed data. We will provide a more detailed explanation of the reasons for model selection in the revised manuscript.

   [1] Li, Y., Zeng, H., Zhang, M., Wu, B., Zhao, Y., Yao, X., Cheng, T., Qin, X., and Wu, F.: A county-level soybean yield prediction framework coupled with XGBoost and multidimensional feature engineering, International Journal of Applied Earth Observation and Geoinformation, 118, 103269, https://doi.org/10.1016/j.jag.2023.103269, 2023.

2) Secondly, **regarding the specific process of model selection**, we adopt a nested cross-validation strategy to optimize the hyperparameters and evaluate the performance of the three models. Specifically, we first divide the samples into a training set and a test set. On the training set, we use 5-fold cross-validation to perform grid search optimization on the model hyperparameters. By traversing different hyperparameter combinations, such as the number of trees and maximum depth of Random Forest, learning rate and number of trees of XGBoost and CatBoost, etc., we find the optimal hyperparameter configuration for each model. Then, we retrain the models using the optimized hyperparameters and evaluate the prediction performance of the models on the test set. The model performance evaluation metrics is coefficient of determination (R^2), which measure the model's ability to explain production variations. We select the model with the highest R^2 on the test set as the optimal model for that agro-ecological zone. It should be emphasized that the above hyperparameter optimization and model evaluation process is carried out independently within each agro-ecological zone to obtain the optimal model for the characteristics of different regions. This strategy helps to improve the model's ability to characterize regional production variation

characteristics and enhance the accuracy of regional production estimation. In the revised manuscript, we will also display the optimal model for each region in a spatialized manner (as shown in **Figure 1**) to intuitively show the advantageous distribution of different models in each agro-ecological zone, aiming to reveal the association between model selection and regional characteristics, and explore the potential relationship between model applicability and regional factors such as climate, soil, topography, and planting systems. It should be noted that this figure is intended to intuitively show the spatial distribution pattern of model selection results. To simplify the illustration, we have not shown national boundaries. For countries without subdivided agro-ecological zones, model selection is performed at the national scale.

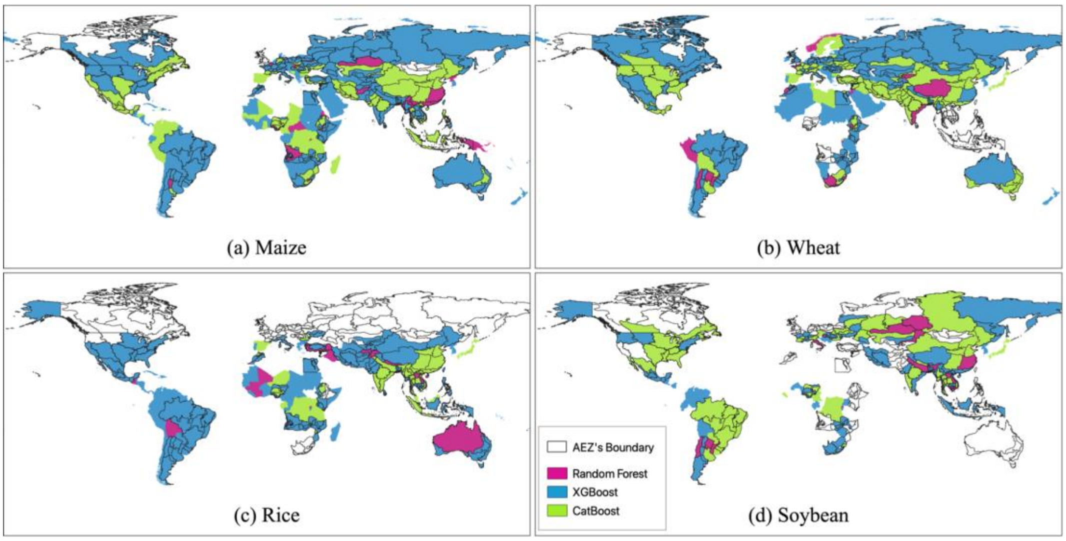

Figure 1. Spatial distribution of optimal model selection results.

3) In addition, if the paper is finally published, we commit to providing complete model training and evaluation code, as well as detailed code documentation, which will help readers fully understand our modeling process and reproduce or improve our methods in other studies.

Through the above supplements and improvements, we believe that the transparency of the model selection section will be greatly enhanced, allowing readers to better understand and evaluate our modeling ideas. Thank you again for your valuable suggestions.

Comment 6:

Data Limitations: While the paper discusses data limitations briefly, a more thorough exploration of potential data limitations, such as inaccuracies in remote sensing data or potential biases, would provide a more comprehensive view.

Response to Comment 6:

We strongly agree with the your viewpoint that a comprehensive discussion of data limitations is valuable for objectively evaluating research results and guiding future research. We will expand and deepen this part in the revised manuscript.

1) **First, regarding the uncertainty of remote sensing data**, we will focus on the following points: Firstly, the quality of remote sensing data is affected by factors such as sensor performance, atmospheric conditions, and surface heterogeneity. In cloudy and rainy areas, clouds and fog will obscure the surface, resulting in missing or decreased quality of data at some spatiotemporal resolutions. In cold high-latitude regions, winter snow cover also affects the extraction accuracy of surface parameters. Secondly, errors may be introduced in the data processing steps such as radiometric calibration, atmospheric correction, and geometric registration, affecting the accuracy of the data. Furthermore, the mixed pixel problem may lead to insufficient representativeness of the extracted surface parameters, especially in regions with severe fragmentation of farmland. These factors may cause systematic biases in remote sensing products in individual regions or time periods, such as overestimating vegetation indices during the growing season.

2) **Secondly, regarding the potential biases in statistical data**, we will focus on the following aspects: First, differences in statistical caliber and standards may lead to insufficient comparability of data between different countries and regions. Second, statistical data may have biases such as underreporting, misreporting, and missing reports, which affect the reliability of the data.

3) Third, **we will provide detailed comparisons between GGCP10 and reference datasets to further explore the strengths and weaknesses of the GGCP10 data** by conducting spatial consistency analyses. For example, in **Figure 2** below, GGCP10 exhibits better spatial transitions compared to SPAM 2010; in **Figure 3**, GGCP10 and USDA survey data show high consistency in high-production counties, while in low- production counties, overestimation or underestimation may exist. In the revised manuscript, we will conduct more detailed analyses to better understand these patterns.

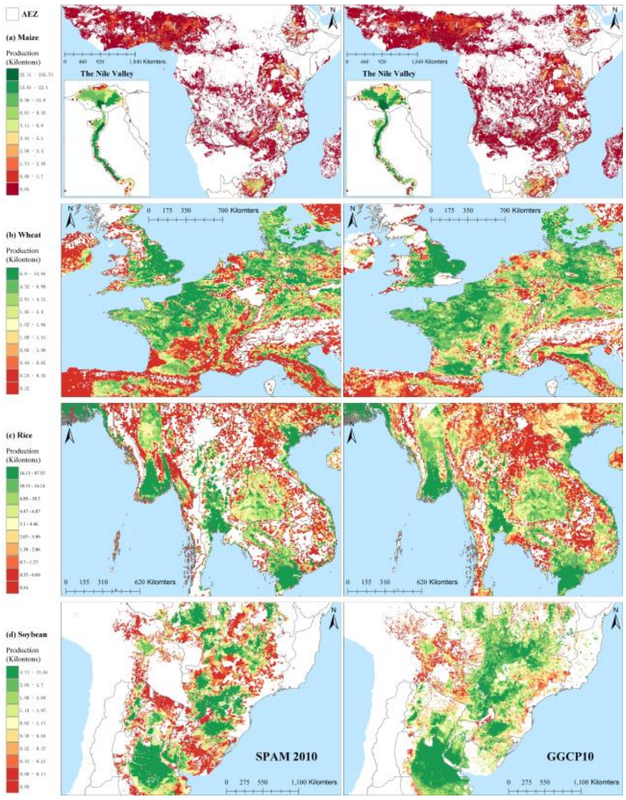

Figure 2. Spatial comparison of crop production between SPAM 2010 and GGCP10 datasets for selected regions: (a) Maize production in Africa; (b) Wheat production in Western Europe; (c) Rice production in Southeast Asia; (d) Soybean production in Brazil and Argentina, South America.

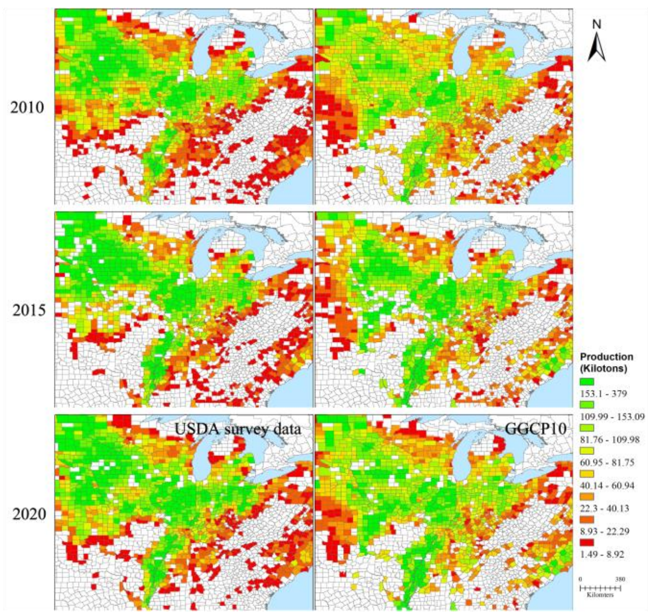

Figure 3. Spatial comparison of Soybean production between USDA survey data and GGCP10 at the county level for the years 2010, 2015, and 2020.

4) **Then, we will also explore the scale effect and spatiotemporal matching issues**. Production statistics are mostly at the administrative unit scale, while model input data are at the pixel scale, and the scale difference between the two may cause errors. The spatiotemporal resolution and boundary definitions of different data products may also not be completely consistent, affecting the accuracy of data matching and fusion. To reduce errors caused by inconsistent administrative boundaries, we consistently used the standard administrative boundary data provided by the FAO in data processing. On the other hand, the spatial resolution of our production data is 10 kilometers, which has a certain scale difference compared to the county-level scale of statistical data. This may lead to a certain averaging effect of the estimated county-level production, resulting in overestimation or underestimation when compared with county-level statistical production data.

5) Finally, we fully understand that **users may encounter situations where our production dataset is not completely consistent with local statistical data** when using it. To address this issue, we will suggest the following processing strategies for users in the revised manuscript: First, it should be recognized that our data may have systematic overestimation or underestimation in individual regions, but overall, it can better reflect the spatial distribution differences of production within the region. Secondly, if users have more reliable regional statistical data, we recommend that users use these data for secondary calibration of our initial estimation results. Specifically, users can calculate the ratio coefficient between our initial estimates and the regional statistical totals, and then use this coefficient to proportionally scale the gridded production data to match the regional statistical totals. This post-processing method not only preserves the spatial distribution information of production revealed by our data but also utilizes local data to improve the accuracy of regional total estimation.

Through the above discussion, we will fully respond to your concerns about data limitations, making the paper's discussion on data quality and applicability more comprehensive and objective. Thank you again for your valuable comments. We sincerely invite you to review the revised manuscript once we submit it and kindly request your continued feedback and suggestions.